Leadership in Sustainable Infrastructure

Leadership en Infrastructures Durables

Vancouver, Canada

May 31 – June 3, 2017/ *Mai 31 – Juin 3, 2017*

REPRESENTATION AND MANAGEMENT OF PROJECT'S KNOWLEDGE - A LINKED DATA APPROACH

Zangeneh, Pouya[1,3], McCabe, Brenda[1], Pearson, Murray[2] and Mason, Nick[2]

[1] University of Toronto, Canada

[2] Hatch Ltd., Canada

[3] p.zangeneh@mail.utoronto.ca

Abstract: Large engineering companies have an abundance of project data in the form of reports and tables. As well, they possess valuable expert knowledge. There is enormous potential to systematically utilise these resources to assist with assessment of risks and estimates. Advances in computer and data sciences have significantly changed our interaction with data. More devices get connected the internet each day. "Sematic Web" has raised as a solution to increase machine readability and interoperability of data using "Linked Data" format and graph data structures. This provides a great environment to capture knowledge using ontology engineering. Overall, these concepts make it possible to create powerful and efficient knowledge bases. This paper presents our path and the latest efforts to create a knowledge management and representation system for engineering projects, using ontology engineering, linked data and semantic web concepts.

1 INTRODUCTION

Megaprojects are the primary mode of development across the engineering sectors of infrastructure, energy, and mining. Including other fields with similar delivery models such as defence, aerospace, and global events, it is estimated that 8%, or US$6 to $9 trillion dollars, of the global gross domestic product (GDP) is spent on megaprojects annually. The infrastructure sector is responsible for nearly half of that figure (Flyvbjerg 2014).

These megaprojects constantly deal with risks and problems due to uncertainty and unforeseen conditions and events. Estimating contingencies for the capital expenditure (CapEx) are among the most important values to inform decision makers. These estimates rely heavily on previous data and engineering experience, and are produced through various engineering procedures. While engineering companies have an abundance of project data to back up these procedures and computational methods have advanced in the last few decades, cost overruns continue to occur despite the care taken to develop these estimates. Large cost overruns have been occurring in last 70 years across all domains around the world (Flyvbjerg et al. 2002).

Cost is typically the most important decision criteria in project selection. That is particularly true for industrial projects with pure economic drivers where other causes of project failure, such as schedule overruns, are eventually translated into costs. For a typical megaproject with CapEx over one billion US dollars, the typical estimating contingency is 5% to 15% of the project's total indicative cost (TIC).

CapEx estimates overrun for a variety of reasons and mechanisms. Of 258 megaprojects across the world, the average cost overrun for roads and rail projects was 20% and 45%, respectively (Flyvbjerg et al. 2002). The mining industry does not perform much better, with average cost overruns estimated between 14%

(Bertisen and Davis 2008) and 22% (Gypton 2002) over the bankable feasibility estimate. An investigation of over 300 industrial megaprojects ranging from US$1 to $20 billion (Merrow 2011) based on key failure criteria of cost, schedule, and production targets, reported up to 65% of the megaprojects failed to meet these key criteria with 40% cost overruns on average. Improving the contingency estimates will provide better financing, better project lifecycle management, and above all, help save private and public resources by identifying better investment opportunities, and strategies to shape projects before sanctioning.

Many algorithms and methods have been proposed to improve CapEx contingencies, including multiple regression analysis (MRA) (Lowe et al. 2006) and influence diagrams (Diekemann et al. 1996). Recent research has incorporated artificial intelligence (AI) algorithms such as artificial neural networks (ANN) (Emsley et al. 2002), fuzzy expert systems (Abdelgawad and Fayek 2010), and case based reasoning (CBR) (An et al. 2007). None, however, have been broadly applied and accepted within the industry, primarily because of a lack of meaningful data. This is especially true since engineering megaprojects are complex systems with many different causal dependencies for which there is never enough data to establish meaningful correlations. The task of creating a reliable model for CapEx requires deep expert knowledge of causalities in addition to the data.

To incorporate explicit expert knowledge, a family of AI algorithms called probabilistic graphical models (PGM) that allow for definition of explicit causalities have been tested. The idea has been effectively utilized by using Bayesian Belief Network (BBN) to deal with schedule contingency (Nasir et al. 2003). Later, BBNs were applied to cost contingency (Khodakarmi and Abdi 2014) as well. Given the effective application of the BBNs and the lack of meaningful data, researchers developed methods to integrate partial probabilistic data with expert knowledge (Das 2004, Zhong and McCabe 2007).

Several root-cause analyses have resulted in often synonymous sets of influential factors, metrics, and measures to explain the reasons behind cost overruns. Poor project phasing was reported in 70% of failed projects (Twigge-Molecey 2003). Statistical analyses have drawn correlations, and often qualitatively elaborated on causations between influential factors and concepts, such as remoteness, team development, and permitting problems across industrial megaprojects and sectors. For instance, the average cost overrun for a variety of contract methods, including lump sum, engineer-procure-construction management (EPC/M), alliancing, and hybrid forms, was reported to be around 15%, 25%, 50% and negative 5%, respectively. Lump sum contracts are considered suboptimal based on cost competitiveness and the tendency to lump too much risk into the cost. The ratio between the winning bid and the first runner up was qualitatively correlated with project failures (Merrow 2011). Therefore, sources of empirical studies and qualitative knowledge exist within the literature to draw causal relationships. Moreover, few industry best practices for quantification of systematic and project-specific risks and their incorporation into cost estimates are proposed in the Total Cost Management (TCM) framework developed and supported by the Association for the Advancement of Cost Engineering (AACE 2012). Building on top of these best practices, certain parametric methods are proposed to incorporate systematic risks through MRA (Hollmann 2016).

The standard estimating procedure in the industry to calculate probability distribution functions for cost or schedule risk is Monte-Carlo simulation. It relies heavily on previous cost data and expert knowledge. Although Monte-Carlo simulation is a valuable tool, it has some significant shortcomings. First, the orthogonality of the parameters assumes each estimated item is an independent variable. This assumption is categorically false for megaprojects, and at best, is countered by certain manually applied correlation factors between the items. Second, the process is reported to be highly subjective, with a normal distribution of contingency estimates across projects using Monte-Carlo simulation, regardless of the principle elements of risk, averaged at 9% with a standard deviation of 4% (Merrow 2011). The use of probabilistic schedule assessment (PSA) methods to assess the failure probability of critical path items may achieve better objectivity (Merrow 2011, Nasir et al. 2003).

System dynamics and discreet event simulation (DES) methods have been proposed - instead of Monte-Carlo simulation - to drive the probability distribution function and to reduce biases and subjectivities in estimating process (Hollmann 2016). Key performance indicators (KPI) have been developed to optimise such simulations by using BBNs to monitor the KPIs and automatically propose corrective responses

(McCabe and AbouRizk 2001). This allows for better resource planning, which is a major project-specific risk for cost overruns. DES is especially useful when the scope definition is complete and the project is ready for construction. However, such high degree of scope definition required for DES is not usually achieved at the end of the bankable feasibility study in megaprojects - when the contingency is set for financing; also, the method does not address the systematic risks and procedural weaknesses.

Hence, although academic research and industry best practices recommend methods to deal with uncertainty and estimates, the longevity and complexity of mega-projects and their procedures still allow for errors, miscalculations, and a wide range of economic, political, and personal biases to impact the outcomes (Flyvbjerg 2009). Addressing this problem in a meaningful way requires dividing it to three main components. First, this problem is due to lack of a comprehensive knowledge base that systematically captures the historical data and knowledge, and maintains corporate memory as it continually expands with new projects. Such knowledge base should be designed to allow efficient and logical deductions and inferences. Second, expanding on previous suggestions (Merrow 2011), the system requires a probabilistic model that integrates data and knowledge to provide objective base-rate probabilistic inference. Third, the system must be implemented within industry environments to create an organizational learning process. In summary, to build on principles of intelligent system design, this system requires, representation, inference, and learning (Koller and Friedman 2013). This paper presents our early efforts to address the first problem.

1.1 Project Data and Knowledge

Industrial megaprojects are not just multidisciplinary in the sense of involving multiple trades and disciplines. They are as well products of multi-agent processes. Although engineering companies typically draw the cost estimate and its contingency, these projects involve multiple owner companies, finance providers, construction contractors, vendors, and in most cases, host governments. Moreover, stakeholders that are impacted by the project can influence the process. These agents not only have different interests; they have different understanding of the project as well as its risks. The causes and reasons of cost overruns can usually be traced back to inefficiencies that involve more than one agent.

The most frequent medium of megaproject data and knowledge is reports and documents. There is no accepted or standardized format for project management data within the industry. The industry’s most recent approach has been the development of Building Information Modeling (BIM) as a systematic design-build modeling process. The necessity for information exchange within BIM processes led to the creation of the Industry Foundation Classes (IFC) as a data model for BIM objects (ISO 16739). Although BIM is extremely powerful, it has not been extended to include project management knowledge, especially outside the architectural, engineering, and construction (AEC) industry. The intersection of AEC with the infrastructure and industrial sectors remains vague, especially when it comes to project management knowledge. Similar to IFC, there have been efforts to create a data model for certain subgroups of industrial projects. Those efforts have resulted in ISO 15926, a standard for the integration of life-cycle data for process plants, and oil and gas production facilities. ISO 15926, for instance, presents a detailed data structure for process and instrumentation diagrams (P&ID) data of process plants. The standard is not extended in any meaningful way to project management data and knowledge. Therefore, this work can as well be considered an extension of BIM or IFC into project knowledge for infrastructure and industrial megaprojects.

There are two main levels of project data: overall project data, and, project lifecycle data. While most efforts in megaprojects research are focused on overall project data, the project lifecycle data appears to be an invaluable - and yet a rare - resource to better understand the projects.

Overall project data considers the project as the main unit of analysis. For example, items like CapEx, schedule, total man-hours, and plant production rates, as well as properties that are not changed during the project like location characteristics. Capturing overall project data requires great diligence in adjusting for currency, inflation, escalation, and location characteristics, for accuracy and provenance. For instance, a billion-dollar project in China at 2010 can not be compared with a billion-dollar project in Canada at 2015 without proper adjustments. A knowledge base should allow for such adjustments and the integration of partial data in an efficient way.

Project lifecycle data, however, takes sub-projects or project phases as the main unit of analysis. The issue with capturing project lifecycle data is that different methods and philosophies of project phasing and planning are employed in different sectors. The two methods used extensively today in approaching projects were originally developed by Construction Industry Institute (CII) (Gibsin and Dumont 1996) and Rand Corporation (Merrow et al. 1981). The phasing philosophy is basically the same for both methods, and consists of dividing the project to four or five phases of conceptual, prefeasibility, feasibility, and execution or their equivalents. CII created the project definition rating index (PDRI) across all sectors of buildings, infrastructure, and industrial projects. Rand Corporation conducted many of the early research in industrial projects and developed the Front-End Loading (FEL) phasing and rating systems. Later, the FEL system was expanded in detail (Merrow, 2011). Most industrial projects are developed on some version of the FEL phasing system in conjunction with AACE best practices for estimation classes. PDRI is often adopted within the infrastructure and building sectors. Aside from the procedures on project phasing and scope definition, few sources have published guidelines on data collection and management in megaprojects for future inference (Hollmann 2016; AACE 2012).

The lack of literature on data driven approaches in estimating and inferences across the megaprojects domain has largely resulted from the confidentiality of these data. High level descriptions of historical projects database design are seldom reported (Musgrove 2008). Most project databases, even within one engineering firm, exist in disconnected data silos. Unfortunately, historical data is often not utilised to any of its potentials and loses relevance with time due to the lack of proper adjustments. A project knowledge base can help facilitate data collection, processing, and utilization across the industry. The unified knowledge base design can happen without compromising the confidentiality of the data instances. Moreover, such medium is increasingly necessary in the current landscape of open public data to increase the effectiveness of the cause.

1.2 Knowledge Base Design

The ability to perform logical deductions is of great benefit to create a project knowledge base for efficient probabilistic inferences. This, as discussed, is due to complexities of megaprojects across different sectors, and the need to use and partition the partial data. Moreover, interoperability and reusability of the data structure and knowledge base is of major importance for any meaningful implementation. An implementation of graph data structure on more than twenty million lessons learned documents in National Aeronautics and Space Administration (NASA) reported a significant increase in utilization of the database by projects staff, as opposed to the previous relational database with SQL query system (Maza 2015). Considering the great potential of logical deduction and inferences through ontology engineering, and interoperability and reusability through linked data and semantic web, the graphical data structure appears suitable for the project knowledge base design.

The literature on the applications of linked data, semantic web, and the ontology engineering around the megaprojects domain is plentiful. A comprehensive review of the field divided these applications based on their content to three general groups, namely, interoperability, linking across domains, and, logical inferences and proof (Pauwels et al. 2017). Although this categorization is a bit vague, it has grouped most applications of ontology engineering in project and construction management and estimation (El-Diraby et al. 2005, El-Diraby 2013, Staub-French et al. 2003) in the “linking across domains” category. What those studies have in common is utilizing ontologies for the process of knowledge transfer within construction management domain. In other words, they deal with pockets of information as their unit of analysis. In contrast, the objective of this research is to create a medium to capture project data and knowledge and to use this information for probabilistic inferences about the projects. Therefore, this research appears to be better aligned with key categories of “interoperability” and “logical inferences” and deductions for expressivity and data manipulation.

2 LINKED DATA AND SEMANTIC WEB

The World Wide Web started in 1989 and has transformed virtually every activity of the modern world. The invention is based on three main technological ideas: uniform resource locators/identifiers (URL/URIs), hypertext markup language (HTML), and, hypertext transfer protocol (HTTP). These three technologies enabled document libraries residing on local server networks among aerospace and defence contractors, academic institutions, and government agencies, to be forged into the modern internet (Leiner et al. 2009). There have been efforts to increase the machine readability of the content on the internet from its early inception. Such efforts first resulted in the 1998 creation of extensible markup language (XML). XML used a series of elements to create and tag the content of a document. These elements facilitated the traceability and to some extent machine readability of the web (Bray et al. 1998). Soon after, XML developers started sharing certain schemas across domains by using domain specific data descriptions. The idea of sharing schemas with URLs suggested the use of URLs for the actual data instances. The semantic web was first introduced in 2001 (Berners-Lee et al. 2001) and gave birth to a simple data format called resource description framework (RDF). RDF is a simple semantic sentence that links a subject through a predicate towards an object, all with universally unique URIs. Figure 1 depicts an RDF triple, their URIs and the links to their universal semantics. The RDF system can store the data, its semantics, and its structure, thereby creating a knowledge base. Later, the potential of RDF format to link between different data endpoints was embraced by the term “Linked Data” (Berners-Lee 2006).

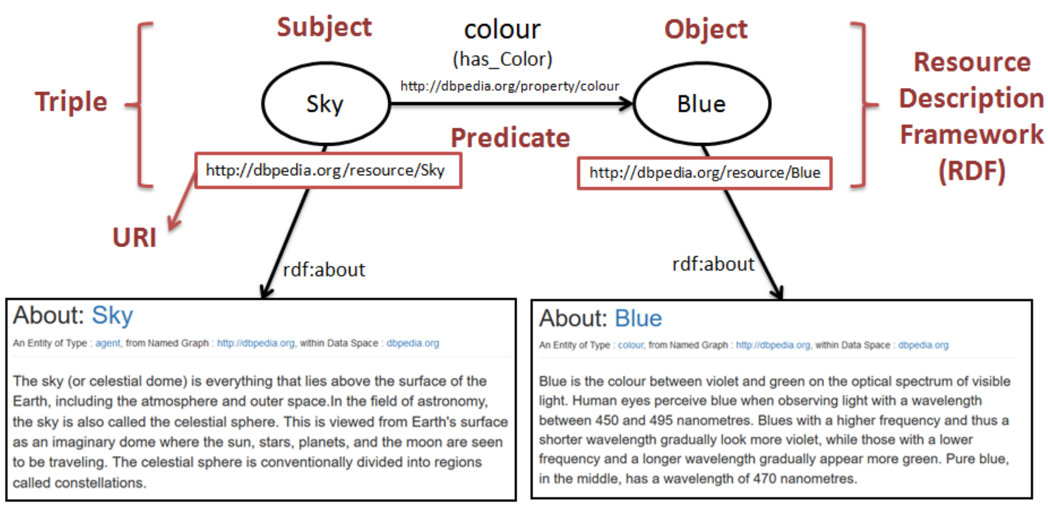

Figure 1: Resource Description Framework (RDF) triple of a simple sentence “Sky is Blue”

2.1 Graph Data Structure

The linked data aspect of the RDF format created a graph of nodes and edges. This structure is powerful in its ability to link data endpoints across the internet. One notable example is DBpedia that created a linked data endpoint of extracted structured information from Wikipedia (Bizer et al. 2009). This created universal semantics for many of the world’s abstract things. Figure 1 shows how objects like sky and blue have unified meanings in DBpedia and can be linked by certain predicates (properties) to their universal semantics. Many applications of Semantic Web and Linked Data are emerging from across the internet to create a shared understanding of the world (e.g. Schema.org (Schema 2017)). This notion contributes to the internet for communication of smart devices, i.e. the internet of things (IoT).

2.2 Ontology Engineering

The use of the term ontology in the context of knowledge representation refers to a shared understanding of a domain of interest. It consists of three main components: explicit description of concepts (classes or concepts), properties of the concepts, and restrictions on properties (axioms or restrictions). The ontology together with a set of individual instances of classes constitute a knowledge base (Uschold and Gruninger 1996; Noy and McGuinness 2001).

The early steps in creating a logical language for the semantic web involved using RDF formats resulted in a language called DAML+OIL (McGuinness et al. 2002). In 2004, the first version of a simple yet powerful logical language based on first order logic called Web Ontology Language (OWL) was introduced by World Wide Web Consortium (W3C 2004). OWL combined the interoperability and machine readability of Linked Data format with the possibility of creating logical representations, and hence, created a powerful, expressive, and interoperable data structure. The application of ontology engineering in the civil engineering have started before OWL, with the first translations of the IFC data model to logical ontologies (Katranuschkov et al. 2003). Completing this translation using OWL opened the door for future research in OWL in the civil engineering industry and the integration of BIM and Semantic Web (Beetz et al. 2005).

3 UNIFORM PROJECT ONTOLOGY

To achieve true interoperability, one must search for prior developed ontologies in the domain of interest. Unfortunately, the authors were not able to find such ontologies with similar purpose to this research. As discussed, most similar efforts are conducted in the realm of facilitating the project work-flow process and not for capturing knowledge. Therefore, it was decided to initiate the uniform project ontology (UP) with the objective of creating a medium for capturing project data and knowledge, on which future probabilistic inferences could be performed.

The task of creating ontology involves identifying the key concepts that describe a megaproject and to devise measures for such concepts by studying the literature and conducting interviews with experts within the field. This is relatively simple for highly tangible concepts like CapEx, but requires great attention for less tangible concepts such as the degree of team development. Such concepts should be dealt with by creating composite measures of several associated factors to achieve a better measure. Given the objective of drawing reliable probabilistic inferences from these measures in the highly complex domain of megaprojects, the reliability and validity of such measures are of great concern. In other words, it is important to make sure those measures record consistent values (reliability), and well represent the actual concept (validity), in all different settings (Singleton et al. 1993). For instance, questions about team development could have different meanings in China versus Canada.

Figure 2 shows the high-level structure of UP ontology with its three classes of Project, Assessment and InfoCard. As discussed, there are various levels of data maturity and availability for megaprojects and there are different assessment methods across industry sectors (e.g. PDRI, FEL). By using “has_Assessment” and “has_InfoCard” properties, an instance of the class Project is related to an instance of its lifecycle and overall project data. The chemical process characteristics and other dimensions like temporal, spatial and knowledge provenance are to be defined in parallel to the principle classes.

The design of an ontology is an iterative process. Considering the scope and objectives are of major importance because increasing expressivity often comes with the price of losing generality (Noy and McGuinness 2001). The UP ontology is designed at high levels to allow for its application and expansion in different megaproject sectors by defining assessment classes. However, the primary goal of this study is to expand this ontology for industrial megaprojects.

Figure 3 shows an adaptation of Natural Resources Canada (NRCan 2016) publicly available database of natural resource projects in UP ontology. The database contains all planned and in-construction natural resource projects across Canada in the mining and energy sectors. Although only high level overall project data was available through NRCan, the flexibility of linked databases allows for partial data to be conveniently partitioned with full data inference. The reflection of the database in Figure 3 is a dimension of the graph database, which is not manipulated as a visualization of the original relational data. The only source of manipulation in the figure is the red circles (projects) that are proportionally enlarged based on the project CapEx. The blue circles represent companies. The network graph structure allows instant realizations from data about the industry and can unfold patterns in any of its different dimensions. such as the distribution of projects with multiple owners, owners with multiple projects, and the companies with projects in both the energy and the mining sectors. For example, it is observable that joint venture projects (projects with multiple owners) are more common in energy sector compared to the mining sector.

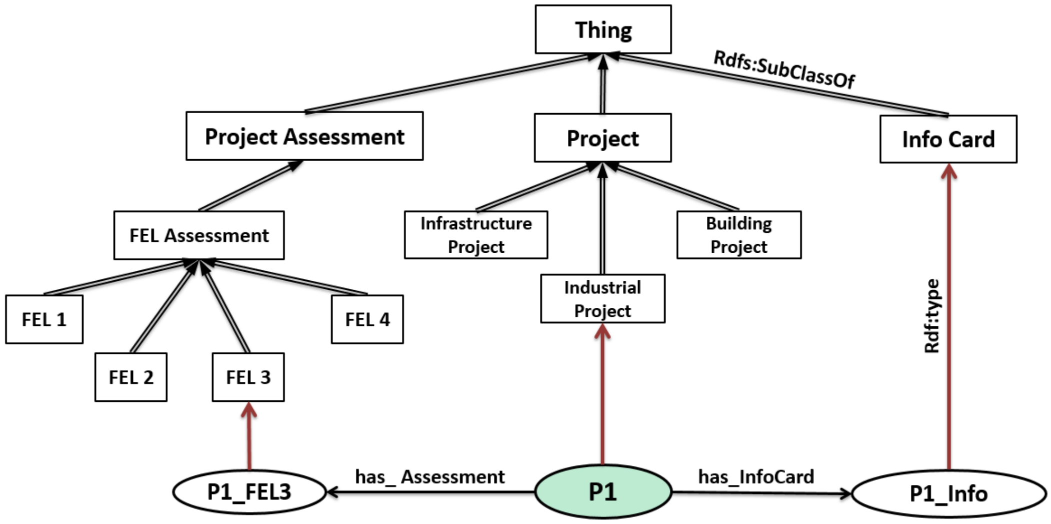

Figure 2: General architecture of UP ontology

The two main characteristics of linked data and semantic web that distinguishes them from the relational database models are observed in Figure 3. First, the database is a graph with semantics. It is interoperable because of the semantics involved and can be seamlessly used by applications. Other linked databases (e.g. DBpedia) can be called in over the internet, without requiring them be hosted in the database (Discoverability). The database can be used as a central infrastructure - fully or partially - for different initiatives (Reusability and/or repackaging of information). For example, Figure *4* is a transfer of BC Hydro energy projects to their corresponding geospatial graph. This sematic rich data infrastructure can be used for different sort of portfolio risk analysis in deferent levels.

Second, the linked database allows convenient logical deductions and inferences. For example, operations can be done on inferred groups of projects with more than one owner company, i.e. joint venture projects. The query results can robustly order the data based on hierarchies and characteristics, (i.e. all projects, versus, all mining projects, versus, all iron and steel projects) which is highly flexible for later probabilistic inferences.

4 FUTURE WORK

In completing the first step of the project, validation of certain measures and their details and specifics for UP ontology is in progress. Also, future work on UP includes the incorporation of temporal and spatial definitions, as well as measures for assigning provenance to knowledge. Expanding on industrial projects, the process definitions shall be added. The next steps involve incorporating the ontology and instances into a probabilistic inference model, namely Bayesian belief networks.

Acknowledgements

We are grateful to many of experts and colleagues at Hatch ltd. for continuing to share their time and knowledge with this project. The project is funded by Hatch ltd. and Natural Sciences and Engineering Research Council (NSERC) of Canada.

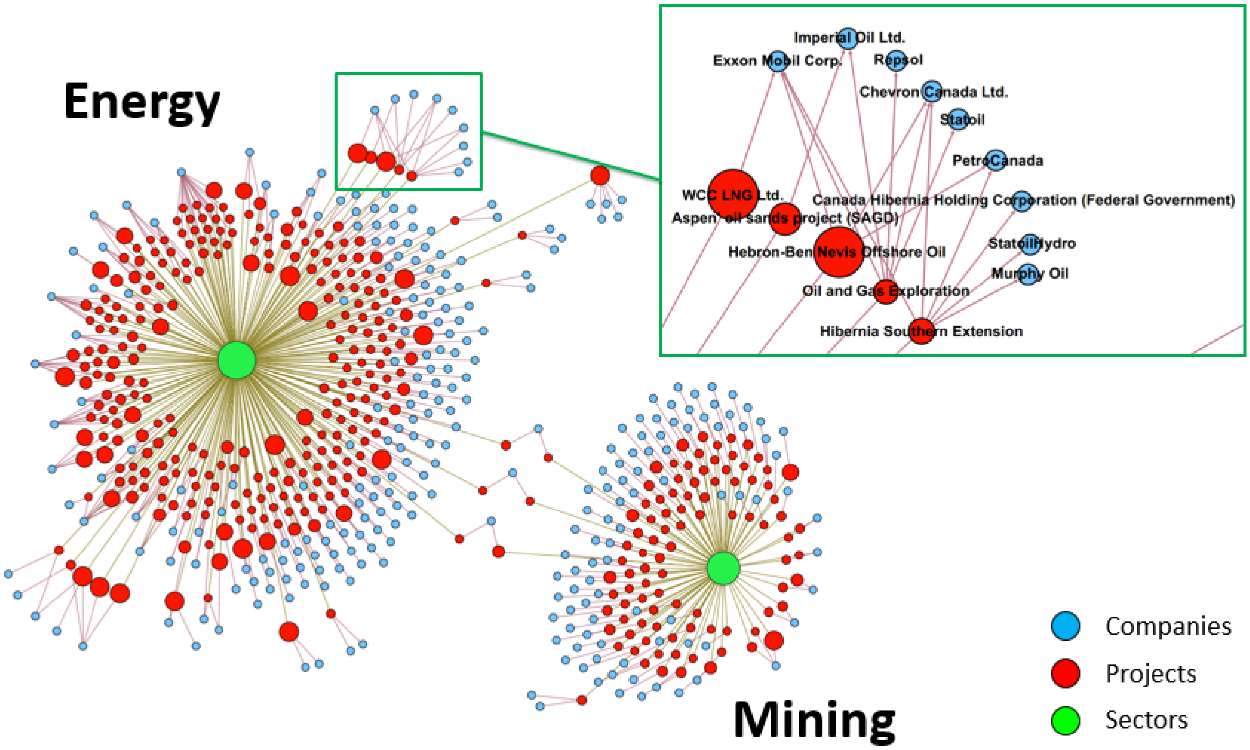

Figure 3: Adaptation of NRC natural resources major project database across Canada as linked data in UP ontology

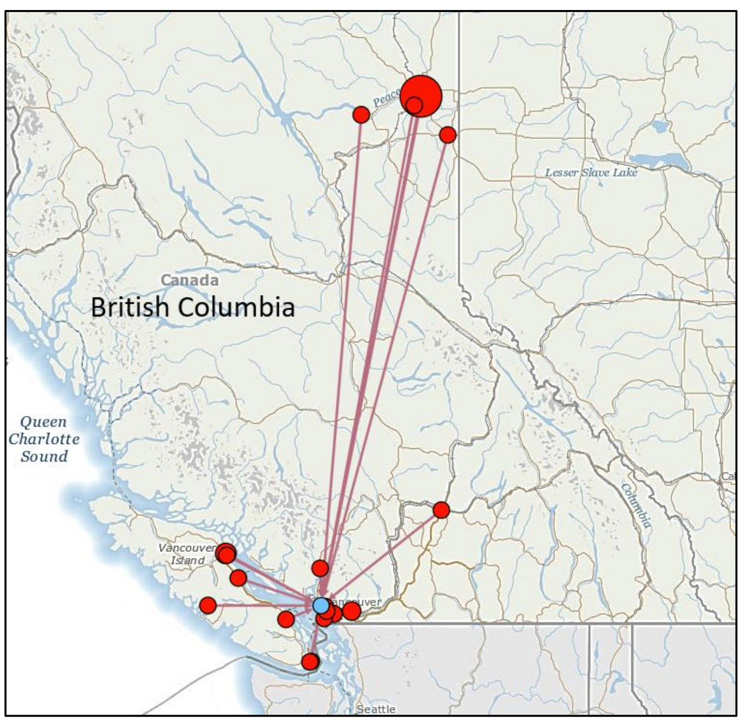

Figure 4: Distribution of BC Hydro capital projects across British Columbia

References

AACE. 2012. *Total Cost Management Framework: An Integrated Approach to Portfolio, Program, and Project Management*. Edited Hollmann, J. K., AACE International.

Abdelgawad, M. and Fayek, A. R. 2010. Risk management in the construction industry using combined fuzzy FMEA and fuzzy AHP. *Journal of Construction Engineering and Management*, **136**(9):1028-1036.

An, S. H., Kim, G. H. and Kang, K. I. 2007. A Case-Based Reasoning Cost Estimating Model Usingexperience by Analytic Hierarchy Process. *Building and Environment*, **42**(7):2573-2579.

Beetz, J., Van Leeuwen, J. P. and De Vries, B. 2005. An Ontology Web Language Notation of the Industry Foundation Classes. *22nd CIB W78 Conference on Information Technology in Construction*, **1**:193-198.

Berners-Lee, T., Hendler, J. and Lassila, O. 2001. The Semantic Web. *Scientific American,* **284** (5): 34–43.

Berners-Lee, T. 2006. Linked Data - Design Issues. <http://www.w3.org/DesignIssues/LinkedData.html> (Feb. 10,2017).

Bertisen, J. and Davis, G. A. 2008. Bias and error in mine project capital cost estimation. *The Engineering Economist*, **53**(2):118-139.

Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R. and Hellmann, S. 2009. DBpedia-A Crystallization Point for the Web of Data. *Web Semantics: science, services and agents on the world wide web*, **7**(3):154-165.

Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., and Yergeau, F. 1998. *Extensible markup language (XML)*, World Wide Web Consortium.

Das, Balaram. 2004. Generating Conditional Probabilities for Bayesian Networks: Easing the Knowledge Acquisition Problem. *arXiv preprint cs/0411034*.

Diekemann, J., David, F., Rhett, M., Keith, M. and Maria, R. 1996. Project cost analysis using influence diagrams. *Project Management Journal*. **3**(4):23-30.

El-Diraby, T., Lima, C. and Feis, B. 2005. Domain Taxonomy for Construction Concepts: Toward a Formal Ontology for Construction Knowledge. *Journal of Computing in Civil Engineering,* **19**(4): 394–406.

El-Diraby, T. 2013. Domain Ontology for Construction Knowledge. *Journal of Construction Engineering and Management*, **139**(7): 768–84.

Emsley, M. W., Lowe, D. J., Duff, A. R., Harding, A. and Hickson, A. 2002. Data Modelling and the Application of a Neural Network Approach to the Prediction of Total Construction Costs. *Construction Management & Economics*, **20**(6):465-472.

Flyvbjerg, B., Holm, M. S. and Buhl, S. 2002. Underestimating Costs in Public Works Projects: Error or lie?. *Journal of the American planning association*, **68**(3), 279-295.

Flyvbjerg, B. 2009. Survival of the Un-Fittest : Why the Worst Infrastructure Gets Built - and What We Can Do About It," *Oxford review of economic policy*, ***25***(3), 344-367.

Flyvbjerg, B. 2014. What You Should Know about Megaprojects and Why: An Overview. *Project Management Journal*, **45**(3): 6–19.

Gibson, G. E. and Dumont, P. R. (1996). *Project Definition Rating Index (PDRI),* Bureau of Engineering Research, University of Texas at Austin.

Gypton, C. 2002. How have we done?. *Engineering and Mining Journal*, **203**(1):40.

Hollmann, J. K. 2016. *Project Risk Quantification : A Practitioner's Guide to Realistic Cost and Schedule Risk Management*. 1st ed., Probabilistic Publishing.

Katranuschkov, P., Alexander G., and Scherer, R. J. 2003. An Ontology Framework to Access IFC Model Data. *Electronic Journal of Information Technology in Construction,* **8**(29): 413-437.

Khodakarami, V. and Abdi, A. 2014. Project Cost Risk Analysis: A Bayesian Networks Approach for Modeling Dependencies between Cost Items. *International Journal of Project Management,* **32**(7):

1233-1245.

Koller, D. and Friedman, N. 2013. *Probabilistic Graphical Models*. MIT Press. Boston, Massachusetts, USA.

Lowe, D. J., Emsley, M. W. and Harding, A. 2006. Predicting construction cost using multiple regression techniques. *Journal of construction engineering and management*, **132**(7):750-758.

Leiner, B.M., Cerf, V.G., Clark, D.D., Kahn, R.E., Kleinrock, L., Lynch, D.C., Postel, J., Roberts, L.G. and Wolff, S. 2009. A Brief History of the Internet. *ACM SIGCOMM Computer Communication Review*, **39**(5):22-31.

Maza, D. 2015. Graphing a Lesson Learned Database for NASA Using Neo4j, R/RStudio & Linkurious. <https://neo4j.com/blog/nasa-lesson-learned-database-using-neo4j-linkurious/> (Feb. 10,2017).

McCabe, B. and AbouRizk, S.M. 2001. Performance Measurement Indices for Simulated Construction Operations. *Canadian Journal of Civil Engineering,* **28**(3):383–93.

McGuinness, D. L., Fikes, R., Hendler, J. and Stein, L. A. (2002). DAML+ OIL: an ontology language for the Semantic Web. *IEEE Intelligent Systems*, ***17***(5):72-80.

Merrow, E. W. 2011. *Industrial Megaprojects : Concepts, Strategies, and Practices for Success*. Wiley.

Merrow, E. W., Phillips K. E. and Myers, C.W. 1981. *Understanding cost growth and performance shortfalls in pioneer process plants*. Rand Corp., Santa Monica, California, USA.

Musgrove, J. G. 2008. If You Build It, They Will Come - Making Project Historical Data Useful. *AACE International 2008 Annual Conference*, AACE.

Nasir, D., McCabe, B. and Hartono, L. 2003. Evaluating risk in construction–schedule model (ERIC–S): construction schedule risk model. *Journal of construction engineering and management*, **129**(5): 518-527.

Noy, N. F. and McGuinness, D. L. 2001. *Ontology Development 101: A Guide to Creating Your First Ontology.* Stanford Knowledge Systems Laboratory.

NRCan. 2016. Free Data - GeoGratis <https://www.nrcan.gc.ca/earth-sciences/geography/topographic-information/free-data-geogratis/11042> (Feb. 10,2017).

Pauwels, P., Zhang, S. and Lee, Y.C. 2017. Semantic Web Technologies in AEC Industry: A Literature Overview. *Automation in Construction,* **73**:145-165.

Shema.org. 2017. <https://www.schema.org/> (Feb. 10,2017).

Singleton Jr., R. A., Straits, B.C., and Straits, M.M. 1993. *Approaches to Social Research.* 5th ed., Oxford University Press.

Staub-French, Sh., Fischer, M., Kunz, J., and Paulson, B. 2003. An Ontology for Relating Features with Activities to Calculate Costs. *Journal of Computing in Civil Engineering*, **17**(4):243-254.

Twigge-Molecey, C. 2003. Knowledge, Technology, and Profit. *5th International Conference of the Copper*, The Metallurgical Society of the Canadian Institute of Mining, Metallurgy and Petroleum, Santiago, Chile, **1**:41-57

Uschold, M. and Gruninger, M. 1996. Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review,* **11**(2):93-136.

W3C. 2004. OWL - Web Ontology Language. <https://www.w3.org/TR/owl-features/> (Feb. 10,2017).

Zhong, T. and McCabe, B. 2007. Developing Complete Conditional Probability Tables from Fractional Data for Bayesian Belief Networks. *Journal of Computing in Civil Engineering*, **21**(4):265-276.